DOCUMENT RESUME

ED 051 261 TM 000 473

AUTHOR	Meyer, Donald L.
TITLE	Bayesian Statistics.
PUB DATE	Feb 71
NOTE	11p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Bayesian Statistics, *Data Analysis, Decision Making, *Mathematical Models, Research Design, *Research Methodology, Research Problems, Sampling, *Statistics

ABSTRACT

Bayesian statistical methodology and its possible uses in the behavioral sciences are discussed in relation to the solution of problems in both the use and teaching of fundamental statistical methods, including confidence intervals, significance tests, and sampling. The Bayesian model explains these statistical methods and offers a consistent model which ties together expectation of results, design of studies, analysis and predictions of future observations. A unique feature of the Bayesian approach is pre-posterior analysis in which one predicts what is likely to happen if a certain sample size is employed and then the expected gain of information is compared with the cost of taking the sample. Finally, this model forces a modeling of the decision making process employed in design and analysis. (PR)

ED051261

Bayesian Statistics

by

Donald L. Meyer
Syracuse University

The purpose of this symposium is to provide a forum for discussion of Bayesian statistical methodology and its possible uses in the behavioral sciences. I will attempt to give a few of the highlights along with some of my thoughts about uses of statistics, in general. If I make strong statements, it is partly because of a defensiveness Bayesians tend to exhibit, but also because I want to generate some discussion and argument from the floor.

TM 000 473

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Presented at AERA National Convention, New York City
February 4-7, 1971

We might date the beginning of the modern era of Bayesian statistics from the publication of L. J. Savage's book, The Foundation of Statistics, in 1954. Even though I was not aware of this book until some years later I think that the seeds of Bayesianism were planted in my mind about that time. This happened when I was exposed to a confidence interval in my first statistics course. I did the usual calculation and ended up with the conclusion that I was .95 confident that the mean of a hypothetical population of I. Q. scores was included in a certain small range of values. I remember that I was genuinely impressed. It didn't really bother me that this was true conditional on the normal model or even that the population was hypothetical. The fact that statistical regularity resulted from randomness excited me then as it still does now. (I really enjoy perusing sampling distributions.) However, I was quickly disturbed and confused when my teacher warned us that the probability that the mean μ was in the interval was either zero or one - that it wasn't .95. I clearly remember the derivation starting from a statement that the probability was .95 that $(\bar{X} - \mu)/\sqrt{s^2/n}$ would be between two t-values and then algebraically obtaining an interval for μ with the final step being the erasure of the P on the blackboard and the substitution of a C. This is, of course, an old story with which you are familiar, but it refers directly to the purpose of doing statistics. We take a sample in order to learn stomething - in order to obtain information. Since we knew a priori that μ was either in or out of a particular interval,

it seems that after sampling we ought to be able to say something more. I interpret the statement of zero or one probability that μ is in the interval as being completely vacuous. Are we then saying that the sample gives us no information?

Some may object and say that at least the confidence interval gives us a likely range of values. But saying "likely" is just another way of saying probable. Nobody needs statistics in order to say "likely". The point is the reporting of some measure of certainty or uncertainty in order to guide us in making decisions- at least mental decisions.

From my experience, most students leave a lecture on confidence intervals feeling confident that the interval calculated in class contains μ, and if statisticians want to play semantic games, well what more could one expect from people constantly dedicated to playing with symbols? Furthermore, I expect and I have some data, that most researchers will give at least even odds, if not 19 to 1, that a .95 confidence interval they obtained in their study does in fact contain the true value of their parameter. Now I submit that we want students to do exactly this. They should mark the test item on our final exam according to the semantic game, but they ought to, in effect, give long odds that a particular interval in their research does contain the true value.

The textbooks generally rated good in our field also do the same thing. They end up with statements that we are .95 confident about a certain interval and then go on to the next topic. I am currently reviewing a textbook copyrighted 1970 which calculates a confidence

interval for a binomial proportion and states that"the probability is .99 that it (the sample) was drawn from a population in which the proportion of events was between .697 and .903." Since the book is non-Bayesian I will have to criticize it on this point, but it does seem sad that the books escaping criticism are those which say nothing meaningful about topics they introduced!

It is fashionable today for many to say they object to formalism of all kinds and that they are data analysts presumably interested in examining reams of output in the hopes of discovering patterns, relationships and what have you. I am in sympathy with this attitude if it is accompanied by the attitude that statistics really is imprecise and that the information content of a particular set of data is known only within fairly broad limits, and furthermore, that this can be communicated only in general terms.

If this is accepted, then our teaching must be revolutionized to the point where topics such as confidence intervals and tests of significance are taught almost as an afterthought for those fortunate enough to have a formal model. In discussing the first course in statistics with the staff of a certain University, they told me that descriptive statistics is taught in only two weeks so that "we get to the important stuff- inference- quickly". I am proud to report that at Syracuse, we spend a full semester on the important stuff - descriptive statistics. Speaking only for myself, I feel that inferential statistics with the accompaning precise statements is not as beneficial to educational researchers as many others seem to imply. Furthermore, the subject is deep and much more difficult

than our textbooks would lead us to believe. I was somewhat [illegible]ked to read in I. R. Savage's book, Statistics: Uncertainty and Behavior that the t-test was not discussed because "...it is of limited theoretical interest and relatively difficult to present and comprehend." I now think I fully appreciate this remark.

I was attracted to Bayesian statistics because it offered a model which could explain things like confidence intervals with logical consistency. The essential idea is that probability is defined as a measure of belief and with a few axions all the usual probability formulae can be derived and, furthermore, the existence of prior distributions for parameters is implied. Data modifies prior distributions by using Baye's theorem, and the resulting posterior distribution summarizes one's current belief about parameters. These concepts - belief and prior distribution - are objected to by non Bayesians. One hears statements that these concepts have no place in science, but the description of how the data was collected and the data itself form the only "objective" facts in a scientific investigation. Everything else is in the nature of an "if-then" relationship. Any further statements are inductive in nature and cannot be separated from belief. Prior distributions are troublesome, but inescapable. The principal difficulty is related to the proper representation of no-knowledge or indifference. The personalistic approach leaves that to the individual while the logical subjectivist searches for general principles which would guide him to a proper choice of starting point for the accumulation of knowledge.

With this background, what can be achieved? For one thing, confi-

dence intervals can now be interpreted. First, under any prior distribution I might have, the prior expectation of the posterior probability included by the $1-\alpha$ interval is $\leq 1-\alpha$. Second, for many models posterior probability that the $1-\alpha$ interval contains the parameter will be approximately $1-\alpha$ for some prior distribution. The practical effect of this is that if one knows this particular prior, and the literature abounds with such results, then one can calculate a confidence interval and readily interpret it in an "if-then" way. If the prior is such and such, then one should bet at 19 to 1 odds on the .95 confidence interval. Furthermore, under wide classes of priors one could bet at least 3 to 1, say. I think that this latter way of thinking is what's behind the data analysis approach mentioned earlier. To a Bayesian everything is interpreted Bayesianly, but informal methods can be interpreted as the search for descriptions or inferences which would hold under wide choices of prior distributions. Consequently, I feel quite comfortable with the data analysis movement.

An example of this is given by I. J. Good (1965). When significance testing a point null hypothesis, if the tail-area probability is .01, then the Bayes factor against the null hypothesis is often in the range of 3 to 1 to 30 to 1 against the null hypothesis. Odds of 20 to 25 to 1 is his best bet. The Bayes factor is the evidence provided by the current data and would be the posterior odds if the prior odds were 1 to 1.

I've mentioned significance tests and other than serving as approximations as just indicated, they are usually a waste of time. We all know that we rarely research true null hypotheses, so there is little point in testing to foregone conclusions. As a rough indication

that the treatment effect is small, they may be useful, but I would rather look at intervals. When we obtain significance which we usually do with large sample sizes, we should go further and inspect the data to see if we have significance in an interesting and useful way. Here, again, our textbooks are remiss. They inevitably warn the student that a significance test doesn't mean practical significance, but then most of the analyses presented is oriented toward the output of one, two or three asterisks. This is, essentially, why non-parametric statistics doesn't interest me a great deal. Many or most of the very clever calculations are pointed toward the single purpose of finding an alpha level.

Another poor feature of significance tests, at least in the way we teach them, is that they are cast in a decision theory mold. We have to accept or reject hypotheses and that is a decision with consequences. Therefore, in order to be rational, we should consider the loss structure. I don't want to argue today whether everything is actually decision theory or not, but the Bayesian can stop at the descriptive level by merely reporting his posterior distribution without accepting or rejecting anything. The results of a significance test, as discussed earlier, can also be interpreted descriptively by a Bayesian.

One point I've never seen in our textbooks about fixed α tests is the fact that one is acting inconsistently if one tests at $\alpha = .05$, say, regardless of sample size. For some small sample size, one could calculate the probability of a beta error for some alternative value to the null. If it is greater than .05, one might justify the choice of

$\alpha = .05$ by saying that the type one error would be the costlier of the two errors. However, for some larger sample size, the beta error for that alternative will be less than .05 so that if $\alpha = .05$ is retained, it must be that the beta error is now the costlier of the two. This implies that as sample size increases, α should be decreased. This would alleviate the "reject with large sample size effect", and in this respect, the tests would resemble Jeffrey's Bayesian tests.

Besides giving explanations of so-called classical statistical methods, the Bayesian approach offers a consistent model which ties together expectation of results, design of studies, analysis and predictions of future observations. For example, suppose one is at the design stage of a proposed study. One question commonly asked is "How large a sample should I take?". Several suggestions or "classical" answers to this question are available, among them being to take a sample large enough to insure a certain small width of a resulting confidence interval or to insure that the power of a statistical test for some alternative to the null will be high enough. These may be fine as far as they go, but the Bayesian solution goes a good deal farther and has a unique feature which we will see.

Briefly, the Bayesian uses his prior expectation to predict what is likely to happen if he takes a certain size sample and he then compares the expected gain of information with the cost of taking the sample. His optimal size sample is the one which has the largest expected net gain.

Before getting into details, it seems obvious that one only has expectation or belief when designing a study. Those who say that

belief has no place in science are making a statement that is just too general.

The Bayesian solution is known as pre- posterior analysis. In a binomial sampling problem, for example, suppose we plan on a sample of size 10. For each possible sample result of zero to ten successes, we do a Bayesian analysis converting our prior distribution to a posterior distribution. We now have eleven possible posterior distributions. Which of these eleven would we like? We might desire a posterior with small variance-especially if the loss function is squared error. Therefore, we could calculate eleven variances - one for each of these possible posterior distributions. It seems reasonable to average the eleven, but what weights should be used? Since our prior distribution for the parameter implies a prior distribution for the number of successes to be observed, we use the prior to give us our weights. We then end up with an expected variance of posterior for that sample size. We can do this for many sample sizes, each time obtaining an expected variance. In general, the larger the sample, the smaller the expected variance, but when cost of sampling is considered, an optimal balance is found.

The unique feature of this approach is that the optimal sample size may turn out to be zero! This means our best action is not to do the study. This solution occurs when our prior beliefs are such that for any economical range of sample sizes the expected sample results would be relatively uninformative.

We make judgements like this when we counsel students or others not to do a study because the anticipated results would be a re-hash

of known results or when results of a pilot study guides us to extensive revisions of plans.

The advantage of the Bayesian model is that it forces those making judgements to bring all assumptions and beliefs out in the open. The resulting gain is that one can study the separate components and their interactions for their effect on the various decisions one makes both in design and analysis. In other words, one can model the process. Isn't this the point of the mathematical exercise in the first place?

References

Good, Irving J. Discussion following "Bayesian interpretation of standard inference statements", Journal of Royal Statistical Society (Series B) 27: 169-92; No. 2., 1965.

Savage, I. Richard. Statistics: Uncertainty and Behavior. Boston: Houghton Mifflin Co., 1968. 344 pp.

Savage, L. J. The Foundations of Statistics. New York: John Wiley and Sons, 1954. 294 pp.